December 2006

The Lord Of The Ratings: Is A Movie's Fate is Influenced by Reviews?

Xiaoquan Zhang
HKUST

Chrysanthos Dellarocas
University of Maryland

THE LORD OF THE RATINGS: HOW A MOVIE'S FATE IS INFLUENCED BY REVIEWS?[1]

Web-based Information Systems and Applications

Xiaoquan (Michael) Zhang
HKUST & MIT Center for Digital Business
zhang@ust.hk

Chrysanthos (Chris) Dellarocas
R. H. Smith School of Business
cdell@rhsmith.umd.edu

Abstract

Third-party reviews play an important role in many contexts in which tangible attributes are insufficient to enable consumers to evaluate products or services. In this paper, we examine the impact of professional and amateur reviews on the box office performance of movies. Using a simple diffusion model, we establish an econometrics framework to control for the interaction between the unobservable quality of movies and a word-of-mouth diffusion process and thereby estimate the residual impact of online amateur reviews on demand. The results indicate the significant influence of the valence measure (star ratings) of online reviews, but their volume measure (propensity to write reviews) is not significant once we control for quality. Furthermore, the analysis suggests that the variance measure (disagreement) of reviews does not play a significant role in the early weeks after a movie opening. The estimated influence of the valence measure implies that a one-point increase in the valence can be associated with a 4–10% increase in box office revenues.

Keywords: Diffusion model, motion picture industry, online reviews, third party reviews, word-of-mouth

Introduction

In the summer of 2004, Steven Spielberg released the movie *The Terminal*, a film with a $75 million production budget and $35 million marketing budget and starring Tom Hank and Catherine Zeta-Jones. During its opening weekend, nearly 3000 theaters showed the movie. Despite these factors that would seem to guarantee a big hit, the movie grossed only $77 million, far less than any predictions had suggested. Notably, the opening gross of *The Terminal* was $19 million, comparable with Tom Hanks's other movies (*e.g. The Road to Perdition*, $22 million; *The Green Mile*, $18 million; *You've Got Mail*, $18 million; *Forest Gump*, $23 million). Thus, the lower total gross of *The Terminal,* less than half the average of his other movies ($157 million), suggests evidence of the effects of bad buzz.

In "experience goods" markets, third-party reviews play an important role in consumer evaluations of products or services (Chen and Xie 2005). When tangible attributes (in the context of movies, these might include the budget, marketing, stars, director, and so forth) are not sufficient to function as signals of the utility gains that might be expected from purchase, consumers likely seek third-party opinions to reduce the risk and uncertainty associated with their consumption (Dowling and Staelin 1994). Restaurants, movies, Broadway shows, books, magazine subscriptions, music recordings, car mechanics, online retailers, television shows, hair styling services, lawn maintenance services, and dentists are but a few examples of such experience goods.

[1] We are grateful to the editor and anonymous reviewers at ICIS 2006 for helpful comments.

In such industries, others' opinions are as important to the sellers as they are to the consumers. However, low-quality producers may be hard to distinguish from high-quality producers of such goods, and if consumers are willing to pay only for expected gains, producers have less incentive to produce high-quality offerings. In return, only low-quality goods get produced and sold for low prices, a problem typical to the "lemons" market (Akerlof 1970).

One way to address the lemon problem is through repeated interactions, but for goods such as movies, books, and shows, repeat purchase is not a feasible option. Others' opinions can come to the rescue in such circumstances. For example, consumers traditionally use *Consumer Reports* (professionals) ratings to learn about the quality of products; more recently, Web sites like BizRate and Epinions maintain scores, based on consumer (amateur) feedback, of most online stores. In an online setting, the limited dimensions of store characteristics and virtually homogeneous nature of the products means higher scores translate into higher perceived quality and thus higher profit for the sellers.

Literature shows that both professional (Litman 1983, Eliashberg and Shugan 1997, Holbrook 1999, Basuroy et al. 2003, Elberse and Eliashberg 2003, among others) and amateur (Godes and Mayzlin 2004, Dellarocas et al. 2005, Chevalier and Mayzlin 2005, for example) reviews play important roles in consumers' decision process, as is summarized in Table 1[2] .

Most of the research listed in Table 1 investigates professional reviews in the movie industry. Amateur reviews have been less popular study topics because they have not been easily observable or measurable until very recently. In addition, the motion picture industry has served as a study context for various reasons. First, movies offer a very representative example of an experience good, because their quality cannot be assessed easily without consumption, so consumers must incur not only the cost of a movie ticket but also the opportunity cost of time. Second, after a movie has been released, word-of-mouth (henceforth, WOM) plays a crucial role in determining its success or failure, which enables researchers to separate the relevant factors into prerelease (e.g. production budget, marketing cost, stars, directors, genre, MPAA rating, professional critic reviews) parameters and post-release (number of theaters for each week, competition with other movies, seasonality impacts, and of course WOM) parameters. Third, the ticket prices of movies are generally constant over time and do not vary across movies, making it easier to avoid complicated demand/supply analyses. Last but not least, movie data sets are easy to obtain, and it is not difficult to compare research results.

With regard to *professional* movie reviews, Eliashberg and Shugan (1997) show that the correlation between reviews and demand may be spurious and conclude that critics' reviews can be used only as "predictors" (instead of "influencers") of total box office grosses.

Regarding to *amateur* reviews, it is generally accepted that the reviews can serve as a *proxy* for WOM, and a model that takes professional or amateur reviews into consideration can provide useful information about the quality of a product. For example, Dellarocas et al. (2005), using a movie revenue forecasting model, show that with only three days of box office and user and critic ratings data, their model achieves forecasting accuracy levels that a previous model (BOXMOD: Sawhney and Eliashberg 1996) required between two to three weeks of box office data to achieve. Godes and Mayzlin (2004) also conclude that online conversations (in Usenet newsgroups) offer an easy and cost-effective opportunity to measure WOM.

[2]Extensive literature on movie revenue forecasting models exists, but due to the scope of this research, we do not review these papers here. Readers may find Sawhney and Eliashberg (1996) and the references therein interesting in this regard.

Table 1. Previous Research Related to Professional and Amateur Reviews

Study	Method	Data	Review	Findings
Litman (1983)	Multiple regression	Movies 1972–1978	Critics	Critics' ratings are significant factors to explain box office revenue
Sawhney and Eliashberg (1996)	Forecasting model, generalized gamma	Movies 1990–1991	Critics	Critics' reviews are positively significant for many adopters
Eliashberg and Shugan (1997)	Correlation analysis	Movies 1991–1992	Critics	Critics' reviews predict box office performance (not influence)
Holbrook (1999)	Multiple regression	Movies Pre-1986	Critics	Ordinary consumers and professional critics emphasize different criteria in the formation of their tastes, but the correlation between popular appeal and expert judgments is positive
Basuroy , Boatwright, and Kamakura (2003)	Diffusion model	Movies 1992–2001	Critics	Valence of critics' reviews affect diffusion pattern of box office
Elberse and Eliashberg (2003)	Demand/supply model	Movies 1999	Critics	Less positive reviews correspond to a higher number of opening screens, but more positive reviews mean more opening revenue
Godes and Mayzlin (2004)	Multiple regression	TV shows 1999–2000	Amateur	Online conversations offer one way to measure word-of-mouth
Reinstein and Snyder (2005)	Difference in difference	Movies 1999	Critics	Critics' influence is smaller than previous studies would suggest but still significant
Dellarocas, Awad, and Zhang (2005)	Diffusion/forecasting model	Movies 2002	Amateur	Online amateur movie ratings can be used as a proxy for word-of-mouth
Chevalier and Mayzlin (2006)	Difference in difference	Books 2003–2004	Amateur	Online amateur book ratings affect consumer purchasing behavior

Despite these findings, the *influence* of professional or amateur reviews/ratings on consumer purchasing behavior has not been well established. One of the major difficulties related to estimating the impact of reviews on box office revenue is the endogeneity problem — movies with higher intrinsic quality tend to have better reviews, so it is hard to determine whether the positive review or the high quality of a movie is responsible for its high demand. Although endogeneity is not an issue for revenue forecasting models, it becomes a serious concern whenever the influence of reviews on revenue is to be implied. In the literature, earlier papers (Litman 1983, Litman and Kohl 1989, Sochay 1994, for example) generally do not consider the endogeneity issue, which means the findings of a positive influence of critics' reviews on box office revenue may be spurious. More recent work has addressed this problem using several approaches. Reinstein and Snyder (2005) and Chevalier and Mayzlin (2005) use *"difference-in-differences*" methods to eliminate fixed effects over time and across different critics (Siskel and Ebert) or Web sites (Amazon and Barnes & Noble). Elberse and Eliashberg (2003) propose a *simultaneous-equations model* to address the simultaneity of audience and exhibitor behavior. Elberse and Anand (2005), in examining the causal relationship between movie advertising and revenue, use *panel data analysis* to eliminate the fixed effects of movie quality. Dubois and Nauges (2006), on the basis of an empirical framework proposed by Levinsohn and Petrin (2003), develop a structural *panel data model* to control for and identify the unobserved quality of wines.

Empirical work on the influence of professional or amateur reviews on movie box office performance is sparse. In the case of professional reviews, Eliashberg and Shugan (1997) find no evidence that reviews influence revenue, but both Eliashberg and Shugan (1997) and West and Broniarczyk (1998) quote a Wall Street Journal survey in which "over a third of Americans seek the advice of critics when selecting a movie"[3] . West and Broniarczyk's empirical study also suggests that consumers respond to critic disagreement. Moreover, movie studios consistently quote favorable critics' reviews in their promotions, in the hope of influencing people to watch the movies. These pieces of evidence cast doubt on the "predictor but not influencer" conclusion. In this paper, we briefly revisit this question with new evidence and discuss the role of professional reviews.

With regard to amateur reviews, Chevalier and Mayzlin (2005) is the most relevant to this discussion. They find that users' book reviews on Amazon.com or BarnesandNoble.com can influence the sales of the reviewed books. In the movie context, online amateur reviews, such as those posted on the Yahoo! movies site or RottenTomato.com, are becoming more and more popular. Although no direct verification exists to suggest more people check Yahoo! movie ratings than have previously, the increasing numbers of reviews written for each movie[4] and movie advertisements on the Web site imply indirect evidence. According to the MPAA (http://mpaa.org, *2005 MPA Market Statistics*), its member companies spent $0.84 million on online advertising in 2005 for an average film, compared with $0.35 million in 2001 and $0.24 million in 2002. Given the growing penetration of the Internet in U.S. households[5] , online amateur reviews likely play increasingly important roles in helping people make more informed decisions about the movies they want to watch.

We investigate whether the influence of these reviews can be identified, paying special attention to the endogeneity problem discussed previously. The identification of the impact of online ratings on product adoption (movie attendance in our case) has important theoretical and practical implications. Word of mouth, as a form of information transmission among consumers, has received relatively little empirical examination in the literature. The availability of online ratings provides us with a way to look into the problem, and is itself an interesting artifact of information. It is still unclear whether consumers include online word of mouth as a factor in their decision making process. If the answer is positive, then an estimate on the magnitude of the impact can be very useful for the practitioners (movie studios, theaters in our context). In this paper, we build a structural model for the word of mouth diffusion process and estimate that one point increase in the average amateur rating on Yahoo! Movies can be associated with 4-10% increase in the movie revenue.

This paper proceeds with the following four sections. The next section describes the data and measures, then we briefly discuss the role of professional reviews. In the following section, we build a diffusion model for WOM, examine the impact of online reviews on demand, and present some robustness checks. Finally, we conclude.

Data and Measures

The data set consists of movie production information, weekly box office information, and Yahoo! Movies professional and amateur reviews of nationally released movies between July 4, 2003, and September 10, 2004. A

---

[3] *Wall Street Journal*, "A 'Thumbs Up' Pulls in the Audience," March 25, 1994, B1.

[4] In 2004, the most popular movie received 97,132 ratings on Yahoo! movies, one order of magnitude more than the most popular movie got (8,754) in 2002. The average number of ratings received in these two years are 9,446.87 and 409.28, respectively. Source: the authors' calculation.

[5] In 2005, the statistics were 70.7 million (63%); in 2001, they were 50.9 million (47%). Source: US Census Bureau, IDC.

few movies are left out of the final sample because they received too few amateur reviews. The final data set comprises 128 movies.

Movie and Box Office Data

For each movie, we collect information about the name, release date, distribution studio, production budget, estimated marketing spending, running time (duration), genre,[6] MPAA rating,[7] and total gross. Weekly box office performance data are also available for these movies. Specifically, we gather the total number of weeks a movie is in the theater, the weekly gross, the number of theaters showing the movie in any week, and the number of competing movies for a movie in any week.

Table 2. Summary Statistics for Movies

Variable	Median	Mean	Std	Min	Max
Budget (million $)	30	39.16	36.32	0.046	200
Marketing (million $)	15	14.8	12.8	0	50
Running time (min.)	102	105.44	22.47	76	200
Opening gross (million$)	9.13	13.49	14.63	0.03	83.85
Total gross (million $)	31.41	46.61	61.23	0.15	377.02
Film life (weeks)	14	14.6	6.3	4	33
# theaters (1st week)	2456	2009.73	1227	2	3703

Movies: 128

The summary statistics in Table 2 are consistent with those published in industry reports (e.g. MPAA theatrical market statistics, *Variety* magazine), which suggests this is a representative sample of movies.

Review Data

We collected professional and amateur reviews for these movies from Yahoo! Movies,[8] which assembles professional reviews from sources like the *Boston Globe*, *Chicago Tribune*, E! Online, *New York Times*, and so forth. According to the Web site, "Yahoo! converts each critic's published rating into a letter grade. If the critic's review does not include a rating, Yahoo! Movies assigns a grade based on an assessment of the review." Amateur reviews are those posted by individual visitors to the Yahoo! Movies Web site. For each review, Yahoo! reports the author's Yahoo ID and date of the review. In addition to writing a textual review, amateur reviewers can rate the movie on the basis of four aspects (story, acting, direction, visual) and provide an overall rating.

[6]Following Yahoo! movie's classification scheme, we designate 12 categories: art/foreign (8 movies), comedy (52), drama (39), romance (26), thriller (17), action/adventure (30), crime/gangster (18), musical/performing (4), kids/family (7), scifi/fantasy (7), western (2), and suspense/horror (15). The numbers in parentheses do not sum to 128 because some movies belong to multiple categories.

[7]There are five rating categories: PG13 (63), PG(17), R (41), NC-17 (1), and Unrated (6)..

[8]http://movies.yahoo.com.

Table 3. Summary Statistics of Reviews

Variable	Correlation with Overall	Mean	Std	Min	Max
Critics	0.5786	7.1953	1.8403	3	11
Overall	1	9.6774	4.0737	1	13
Story	0.9478	9.5739	3.9918	1	13
Acting	0.9185	9.8969	3.7795	1	13
Direction	0.9445	9.5937	3.9273	1	13
Visual	0.8968	10.0876	3.7477	1	13
Total reviews per film	0.3711	1435	2373	27	18269

Table 3 provides the summary statistics of the reviews. For all the scores, we convert letter grades to numerical values, such that an A+ score corresponds to a numerical score of 13 and the lowest possible score,[9] an F, corresponds to a 1. The average professional score is considerably lower than that of the amateur reviews, such that their average grade is a C+ (7.20), whereas that of amateurs is a B+ (9.68). However, the high correlation between them (0.5786) indicates that, to some extent, the "expert judgments" and the "public appeal" agree to each other (Holbrook 1999). In addition, the professionals are more consistent than the amateurs; this is reflected by the much smaller standard deviation for the scores. Professionals also are less likely to give extreme scores like A+ or F, but because many of the grades are assigned subjectively by Yahoo!, this may reflect Yahoo!'s own conservativeness. The scores for the four specific aspects of movies are highly consistent with the overall score, as evidenced by their high correlations. It is interesting to note that amateur reviewers tend to be generally satisfied with the visual and acting aspects and are relatively more picky about the story and direction. There are considerable variations (standard deviation twice as large as the mean) in the total reviews per film, such that the most reviewed movie (*The Passion of the Christ*) received 18,269 reviews, whereas the least (*Buffalo Soldiers*) got only 27 reviews. There is a positive correlation between the valence of reviews and the propensity to review; that is, better movies attract more reviews.

In our subsequent analysis, we match the amateur reviews to the weekly box office data and calculate a cumulative average score for each movie i in week t by summing the overall scores submitted for movie i before week t and dividing by the number of reviews. Note that this score is the online rating that visitors to the Web site see when they check on movie i during week t. In addition to this valence measure, we calculate volume and variation measures. The volume of movie i for week t is measured as the number of reviews posted before week t for movie i, and the variation is a measure of the level of consensus among consumers. We use common dispersion measures (variance, Gini coefficient, coefficient of variation) to evaluate this level of agreement.

The Influence of Professionals

Since the genesis of movies, as early as the silent film era, film criticism has been recognized as an art. After the 1940s, movie criticism became a profession as well. In his new book *American Movie Critics*, Phillip Lopate (2005) calls the period from 1950s to the 1970s "the golden age of movie criticism." Film critics like Pauline Kael and Andrew Sarris were read widely, and their reviews had the power to cause public stirs. Later, the television program

[9]In Yahoo! Movies, it is not possible to give scores of E-, E, or E+, so from F to A+, there are 13 levels.

Siskel and Ebert (changed to *Ebert & Roeper* in 1999) made its reviewer stars household names and has been vastly successful for two decades. It is hard to imagine that these reviewers could have been so popular for so long without having an impact on the box office performance of movies. Film studios obviously believe that good reviews from critics help attract viewers, as they routinely quote terms such as "spectacular," "thrilling," and "two thumbs up" in their movie advertising blurbs. To get better reviews, studios have been known to "bribe" reviewers by flying them first-class to the studios for a weekend and offering generous gifts. However, if they cannot easily predict reviewers' reactions, studios often refuse to offer critics previews to avoid getting slashed (e.g. *The Avenger*).

Eliashberg and Shugan (1997) study two possible critics' effects: influence and prediction. Whereas influence implies an impact, prediction suggests that critics' reviews are merely leading indicators that have no significant impact on actual box office revenues. They regress weekly box office revenues on measures of critics' reviews and find that box office revenues after the fifth week are correlated more strongly with critics' reviews, suggesting in favor of a "predictor" story. Reinstein and Snyder (2005) attempt to exploit the timing of Ebert's and Siskel's reviews to distinguish the influence and prediction effects and find weak evidence of an influence effect.

In this section, we replicate Eliashberg and Shugan's (henceforth, ES) study with a new data set and compare the results with theirs. In Table 4, we report summary statistics for movies and critics' reviews, following the format of Table 2 in ES.

Table 4. Motion Pictures and Critics' Summary Statistics

Variable	Median	Mean	Std
Film life in weeks	14	14.6	6.3
Screens (1st week)	2456	2009.73	1227
Box office (1st week)	9,130,000	13,490,317	14,633,000
Cumulative box office	31,410,000	46,613,359.40	61,234,610
Total reviews per film	13	11.89	3.18
Percentage of positive reviews	46.60%	47.30%	0.299
Percentage of negative reviews	21.40%	29.10%	0.278

ES's data set contains 56 movies released in 1991–1992. As Table 4 shows, 10 years later, the median film life in weeks is slightly shorter. The number of screens more than doubled, from 1,122 to 2,456, and the first week's box office revenue almost tripled, from $3.6 million to $9 million. However, total gross circa 2004 ($31 million) is only slightly higher than that in 1991 ($28 million), which appears to indicate that the audience has shifted to the earlier weeks of movies' life cycles. Previously, ES read reviews and decided whether each was positive or negative; for our data set, we simply classify scores higher than B as positive and those lower than C as negative. With this method, the distribution of positive and negative reviews is highly consistent with those in ES.

Table 5. Regression Results for Percentage of Positive Reviews

Week	Multiple R^2 (Adj. R2)	Percentage Positive (Std. Coeff.)	t-Statistic (p-value)	Total Number of Reviews (Std. Coeff.)	t-Statistic (p-value)	Total Number of Screens (Std. Coeff.)	t-Statistic (p-value)	F-Ratio (p-value)
1	0.4958	0.27001	3.83	0.17208	2.5	0.71214	10.62	39.98
	(0.4834)		(0.0002)		(0.0139)		(<.0001)	(<.0001)
2	0.4986	0.31854	4.55	0.18625	2.71	0.68401	10.31	40.44
	(0.4863)		(<.0001)		(0.0076)		(<.0001)	(<.0001)
3	0.5863	0.26788	4.25	0.10831	1.72	0.71879	12.21	56.68
	(0.5759)		(<.0001)		(0.0881)		(<.0001)	(<.0001)
4	0.5801	0.23456	3.65	0.09262	1.45	0.66218	10.99	54.81
	(0.5696)		(0.0004)		(0.1498)		(<.0001)	(<.0001)
5	0.6587	0.13678	2.31	0.12729	2.21	0.72746	13	75.92
	(0.6501)		(0.0224)		(0.0291)		(<.0001)	(<.0001)
6	0.7083	0.13621	2.43	0.136	2.53	0.74462	13.83	92.29
	(0.7007)		(0.0165)		(0.0127)		(<.0001)	(<.0001)
7	0.7128	0.08549	1.45	0.12279	2.22	0.7687	13.49	87.71
	(0.7047)		(0.1496)		(0.0286)		(<.0001)	(<.0001)
8	0.5852	0.05469	0.74	0.17923	2.51	0.67684	9.39	43.74
	(0.5719)		(0.4616)		(0.0139)		(<.0001)	(<.0001)
All	0.4329	0.06839	3.38	0.05485	2.71	0.66102	35.02	409.68
	(0.4318)		(0.0007)		(0.0068)		(<.0001)	(<.0001)

Table 5 shows the result to be compared with Table 4 in ES. The percentage of positive reviews represents a significant variable from the first week, though its significance tapers away slowly to the end of the sixth week. The percentage of positive reviews is most influential in the second week. If the influencer story is true, it may indicate that during the first week, more people who tend to love the movie goes, but in the second week, people who are indifferent will be more influenced by the critics. Other explanations are possible; for example, the critics are merely predictors, but WOM after the first weekend drive people to the cinema in the second week. This scenario is quite possible because comparing the two datasets, we showed above that people tend to watch movies in earlier weeks now than in the early 1990s. The decreasing correlation between critics' review and weekly box office revenues over time supports ES's claim that consumers best remember a review just after they have first read it.

Other specifications (not reported) can compare our results with those of ES as well. For example, the percentage of negative reviews and the valence of average critics' reviews might be included as independent variables. The results are similar --- measures of critics' reviews are highly correlated with the box office revenues starting from the first week. Care must be taken in interpreting this result though; it is not enough to support the influencer claim. Rather, the early significance of critics' reviews may be a result of the audience's shift to earlier weeks or simply because online and offline WOM gets communicated faster and easier than in the past. That is, our analysis may weaken ES's result but does not offer definite evidence to reject or accept it. Further analysis therefore is needed to obtain a clearer understanding of the role of critics.

The Influence of Amateurs

Suppose a movie's quality enters the revenue forecasting model linearly and can be captured by a score (or vector of scores), then a traditional forecasting model can be augmented with such a measure to provide consistent estimates of the impact of both professional and amateur reviews. If quality were controlled for, researchers could examine whether a higher review score induces higher demand and would not be haunted by the endogeneity problem. However, this linearity assumption implies that, when budget, marketing, reviews, seasonality, and other observable characteristics are controlled for, two movies with the same quality score should bring in same amount of total revenue. Because a movie's quality does not change over time, the two movies' weekly box office revenues also should be equal each week, and the weekly revenue trajectories should be the same. This result does not sound plausible. First, a five-star blockbuster movie might have a totally different revenue trajectory than a five-star sleeper movie. The blockbuster movie might be 100 times more costly to produce than the sleeper movie, but that does not mean the total gross will also be 100 times different, nor that the pattern of revenue will follow the same shape. Second and more important, a movie is a social product, and its quality is subjective and depends highly on the "eye of the beholder". A five-star movie to one person might be a one-star movie to his or her neighbor, and this difference does not mean their tastes always are negatively correlated; both consumers might give five stars to yet another movie.

Therefore, the quality measure must enter the equation in a non-linear way. It also may interact with some observables or non-observables. There does not seem to be a good way to write down a meaningful equation — even if we can write such an equation, it remains unidentifiable because the quality is unobservable.

In a famous Harper's Magazine interview,[10] Mark Gill, senior VP of publicity and promotion for Columbia Pictures, said: "It doesn't matter if the movie doesn't deliver. If you can create the impression that the movie delivers, you're fine. That's the difference between *playability* and *marketability*. When you've got playability, you show it to the critics. If you've got marketability, there's enough there to cheat it or stretch it and make people believe that it really does deliver." This distinction offers one way to think about the quality of a movie beyond a uni-dimensional approach. Because our objective in this section is to estimate the influence of amateur reviews on box office revenue, we focus on the "playability" of a movie and largely take opening weekend revenue as given, which depends heavily on the cast, marketing, production, and so forth (i.e., marketability). Overtime, playability will be reflected in consumers' responses to the movie, and consumers' responses constitute WOM about that movie.

To understand this reasoning, consider two movies with the same prerelease parameters (i.e., same marketability measures). If the postrelease weekly number of theaters, competition, and seasonal variations are controlled, the different composition of the audience (consumer mix) for each movie should lead to different adoption patterns [11]. Each week, WOM gets reinforced because when people talk about the movie, they will induce the adoption or non-adoption of other people with similar tastes. This underlying process of WOM getting passed on can be modeled as a diffusion, whose pattern reflects the quality of a movie. A better movie will attract more consumers, and their positive WOM will positively influence the adoption positively, until the point of market saturation. As the preceding discussion indicates, WOM is inseparable from the playability of a movie's quality, and the revenue trajectory may signify a movie's quality through the WOM diffusion process.

WOM Diffusion Model

Suppose in week t the box office revenue of a movie is R_t, and because the ticket prices in different cinemas are essentially the same, R_t can be used to calculate the number of people, A_t, going to the cinema to view the movie in week t. Hence, $A_t = f(R_t) \doteq \frac{R_t}{p}$, where p is the ticket price. Among these consumers, a proportion $\theta(q)<1$ will spread

---

[10] Harper's Magazine, "The New Auteurs; Motion Picture Marketing," June, 1993. http://www.ibiblio.org/pub/electronic-publications/stay-free/ml/readings/new_auteurs.html

[11] Of course, for typical blockbuster movies like *Star Wars* and *The Lord of the Rings*, the composition of the audience may be quite similar (and thus have similar decay patterns).

positive comments (whether off- or online), where q is a combination of the unobserved quality of the movie and the other time-invariant forces correlated with the quality (e.g. production budget, marketing, star power, critics reviews). There are also naysayers, let $\phi(q)<1$ represent this portion of the consumers, and the rest $1-\theta(q)-\phi(q)<1$ say nothing. Together with other forces, ξ, unrelated to q (but constant over time), θ and ϕ determines the consumers who view the movie in week t+1. Thus,

$$A_{t+1} = g[\theta(q), \phi(q), \xi]A_t = g(q, \xi)A_t \tag{1}$$

Note that equation (1) assumes the number of consumers in week t+1 is solely determined by the number of consumers in week t and the WOM process, we will relax this assumption below.

Equation (1) can be used to calculate the change in revenue over time,

$$\Delta R_t = R_{t+1} - R_t = p \cdot A_{t+1} - p \cdot A_t = pA_t[g(q,\xi)-1] = [g(q,\xi)-1]R_t \tag{2}$$

Rewriting equation (2) in continuous form and defining $\lambda \equiv g(q,\xi)-1$, we can get:

$$\frac{d}{dt}R_t = \lambda R_t \tag{3}$$

where the parameter λ measures the playability (WOM diffusion) of a movie.

Note also that we use $g(.)$ in equation (1) to capture the complex relationship among intrinsic quality, other time-invariant forces, and the WOM process. Although we cannot directly observe θ and ϕ, they depend on the quality[12] q, and they should satisfy $d\theta/dq > 0$, and $d\phi/dq < 0$. It is also reasonable to assume, from (1), that $\partial g(\theta)/\partial\theta > 0$, and $\partial g(\phi)/\partial\phi < 0$; thus, $\frac{d\lambda}{dq} = \frac{dg(\theta,\phi,\xi)}{dq} = \frac{\partial g}{\partial\theta}\frac{d\theta}{dq} + \frac{\partial g}{\partial\phi}\frac{d\phi}{dq} > 0$. Therefore λ, derived from a diffusion process, is an increasing function of q, and can be regarded as a measure of quality. The parameter ξ is a time-invariant variable that captures forces unrelated to the quality of the movie, such as unemployment rates, the availability of a public forum to discuss the movie, and so forth. It does not affect the relationship between λ and q because, by definition, $d\xi/dq = 0$.

The solution of (3) is:

$$R_t = R_0 e^{\lambda t} \tag{4}$$

where R_0 is the initial condition, and in this context, it has a perfect interpretation as the opening week's gross. Equation (4) implies that:

$$R_t = R_{t-1} e^{\lambda} \tag{5}$$

Starting from a WOM diffusion formulation, we derive an exponential model[13] for movie revenues that parsimoniously captures a movie's quality as a decay parameter.

---

[12]The parameters θ and ϕ should be movie specific, but because we are examining one movie, we suppress the subscript i in the equations.

[13]Exponential decay models are widely used in similar modeling situations. In similar contexts, Jedidi *et al.* (1998) use an exponential decay model of market share to identify four clusters of movies, and Eliashberg *et al.* (2000) use one to describe exposure to WOM conversations. Einav (forthcoming) uses exponential decay to control for factors other than seasonality. In the same spirit, a generalized gamma model can be used to model the diffusion process(Sawhney and Eliashberg 1996, Ainslie, Dreze and Zufryden 2005). More elaborated diffusion models also can be found in the literature, such as in Dellarocas *et al.* (2005).

Rearranging (5), and adding the terms of amateur review variables, control variables, and an error term, we obtain the empirical model:

$$log\,\frac{R_t}{R_{t-1}} = \lambda + \eta t + \sum_{j=1}^{J} \beta_j x_{jt} + \sum_{k=1}^{K} \gamma_k y_{kt} + \varepsilon_t \tag{6}$$

where λ is the movie-specific fixed-effect quality measure (also a measure of WOM diffusion), η is a time decay factor that is independent of the movie quality, $x_t=(x_{1t},...,x_{Jt})$ is a vector of J independent WOM variables calculated for the t-th period, $y_t=(y_{1t},...,y_{Kt})$ is a vector of K variables for period t that controls for competition and supply, $\beta=(\beta_1,...,\beta_J)$ are the parameters to be estimated to indicate the impact of WOM, and $\gamma=(\gamma_1,...,\gamma_K)$ are parameter estimates of the control variables.

Note that in the specification equation (6), time decay factor η allows the examination of possible changes in the composition of the audience over time, thus the implicit assumption of constant θ and ϕ in equation (1) is relaxed. Two forces may drive the sign of η. If the aficionados see the movie first, followed by the indifferents, there should be a negative η. If, however, people who love the movie induce similar-minded people to go in later weeks (and thus reinforce positive WOM), η should be positive.

The form of (6) is much more flexible than (5), too. Not only the assumptions for constant θ and ϕ are relaxed, the addition of WOM and control variables allows for examination of the post-release forces in contributing to the changes in revenues over time.

When a movie's revenue decreases over time, the right-hand side of the equation should be negative. Because larger absolute values of λ indicate faster decay, a movie with higher quality should have smaller absolute values of λ. For a sleeper movie, the revenue in some weeks can be larger than that of the previous weeks, and model (6) can accommodate this possibility with positive estimate of the right-hand side.

As stated previously, the WOM effect λ can be regarded as the playability aspect of the quality of a movie, and if online amateur reviews influence people's decisions, this effect should also be captured in λ. The purpose of including the variables x is to estimate the *residual impact* of the online reviews, and the estimate of β then captures the residual influence of online amateur reviews. Included in the vector of x, we have variables to measure the (a) valence, (b) variance, and (c) volume of WOM. To illustrate, consider online amateur reviews as one more component of the consumer WOM process. If it works like other means of WOM, λ will capture the influence. Then, introducing the independent variables of online amateur reviews in the model can provide an estimation of the effects of these reviews on revenue only when changes from week t-1 to t cannot be captured by the WOM decay parameter, which this is exactly what we set out to identify as the residual impact of online amateur reviews on box office revenue.

This formulation implicitly (and critically) assumes that, controlling for observables, the decay pattern does not systematically vary with idiosyncratic variations in WOM[14] . We test the validity of this assumption in section 4.2. With this assumption, we can safely use the parameter estimate β to examine the influence of online reviews on the weekly box office performance that is not explained by the general WOM process.

To avoid the potential problem of heteroscedasticity, we conduct White's general test (White 1980) and the Breusch-Pagan test (Breusch and Pagan 1979). Whereas White's test ($p>\chi^2$=0.0412, DF=5) rejects the null hypothesis of no heteroscedasticity at the $p<0.05$ level, the Breusch-Pagan test ($p>\chi^2$=0.4875, DF=2) cannot reject the null hypothesis. Estimates of FGLS (Feasible Generalized Least Squares) have almost the same standard error values as those of the ordinary least squares (OLS) and no significance level changes, so OLS standard errors are reported in Table 6. Four specifications (pure decay, WOM valence, WOM variance, and all previous variables plus controls) are estimated. In the pure decay model, the left-hand side variable is regressed on a constant and a variable that indicates the number of weeks since release. Each other model adds one or more variables to the specification.

[14]Note that this is not an unreasonable assumption because the quality of a movie does not change over time.

The WOM valence is a cumulative average rating of what consumers view at Yahoo! movies before weekend *t*. The WOM variance is measured by the inverse of the coefficient of variation (CV)[15] of the averaged amateur ratings[16]. Following the literature (Zufryden 1996, Zufryden 2000, Jedidi, Krider and Weinberg 1998, Ainslie, Dreze and Zufryden 2005), competition is measured by the number of new movie releases in week *t*. We also include the number of theaters showing the movie in week *t* to control for the possible influence of capacity constraints on the supply of movies. To control for seasonality, we mark the holiday weekends (Martin Luther King Day, President's Day, Memorial Day, Labor Day) as anchors, then calculate the three-year (from January 7, 2000 to December 27, 2002) average total revenue for each week. To create an index for each week, we normalize the average total revenues, such that the lowest grossing week has a score of 1, and a week with 1.5 times revenue earns a score of 1.5. The WOM volume, a measure of how many people make the effort to write reviews, is not significant in any of the specifications.

Table 6. Residual Impact of Reviews on Box Office Revenue

	Pure Decay		WOM Valence		WOM Variance		Controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Movie quality	-0.1882**	-0.5402***	-0.7083***	-0.9107***	-0.4892***	-0.7094 ***	-1.3525**	-0.9810***
	(0.0227)	(0.0455)	(0.1336)	(0.0839)	(0.1539)	(0.0970)	(0.3484)	(0.1842)
Time decay	-0.0909***	0.0096	-0.0926***	0.0079	-0.0942***	0.0054	-0.1299***	-0.0065
	(0.0249)	(0.0071)	(0.0243)	(0.0070)	(0.0241)	(0.0070)	(0.0258)	(0.0091)
WOM valence			0.0589***	0.0422***	0.0511***	0.0347***	0.0961***	0.0434***
			(0.0112)	(0.0081)	(0.0115)	(0.0082)	(0.0267)	(0.0147)
WOM variance					-0.3561***	-0.3088***	0.1809	-0.1913*
					(0.1267)	(0.0760)	(0.2163)	(0.1032)
Competition							-0.0124***	-0.0364**
							(0.0254)	(0.0184)
# Theaters							-0.0001***	-0.00006***
							(0.0000)	(0.0000)
Seasonality							0.3014 ***	0.1984***
							(0.0000)	(0.0000)
R^2	0.0255	0.0017	0.0754	0.0265	0.0896	0.0412	0.2001	0.0727
F-Test	13.29	1.82	20.66	14.58	16.6	15.34	18.39	11.96
Pr>F	0.0003	0.1772	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Note: Standard errors appear in parentheses.

[15]WOM variance=1/*CV*= $1/\mu/\sigma = \sigma/\mu$, where σ is the standard deviation, and μ is the mean. The CV is widely used as a measure of reliability, volatility, or risk. Here, it serves as a measure of the level of agreement among consumers about the quality of a movie. This measure will be greater for more controversial movies.

[16]The CV is a dimensionless number, in that it has no unit and thus allows comparison with numbers that have significantly different mean values. This is approach is similar to the use of entropy by Godes and Mayzlin (2004) to measure the dispersion of online conversations while controlling for the total volume of posts. The use of other dispersion measures, such as Gini coefficient and standard deviation, does not qualitatively change the results.

The odd numbered columns contain the estimates obtained with five weeks of data for each movie, and the even numbered columns report those obtained with ten weeks of data. Because the table included pooled result for blockbuster and sleeper movies, the movie quality parameter λ changes considerably across specifications. Note that all the five-week estimates of λ are smaller in absolute value than their ten-week counterparts, because the sleeper movies have much greater λ values in earlier than in later weeks. In all columns, λ's are negative and significant. Over a ten-week period, an average movie's quality follows an exponential decay rate of -0.9810, which translates to a reduction of 62.5% in revenue every week. According to the preceding discussion, if a movie's decay rate is less than -0.9810 in absolute value, it is a good movie. When the controls for competition, number of theaters and seasonality are added, the parameter estimates of λ and WOM valence increase. The cumulative average score, valence of WOM, entails a significant force with a positive impact on revenue. In various specifications, the parameter estimates are robust, and they all take values from 0.04 (ten-week models) to 0.09 (five-week model with controls). Therefore, for every one-point increase in the score (e.g. from B+ to A-) in a week, the revenue in the next week will be increased from 4.4% to 10%. In columns (5) and (6) WOM variance seems to be very influential in determining the overall decay rate of movies, but once the control variables are added, it is no longer very significant. The control variable "number of theaters" is significant in the model, but its magnitude is too small for it to be important. As expected, seasonality is an important force, in that movies in "good" weeks usually earn much more than they do in "bad" weeks[17] . When the control variables are added to the model, the five-week variance measure is no longer significant, which likely suggests that in the early release stages of the movies, disagreement in tastes may not be as important a factor in influencing the box office revenues. The estimates for change in consumer mix, η, are all significant in the five-week models but not in the ten-week models, which implies that the aficionado effect is more pronounced in earlier weeks.

Robustness Checks

To verify the robustness of the results, we examine two extensions. The first extension verifies that the online rating scores do not merely capture the idiosyncratic weekly variations in offline WOM. Suppose that, due to television advertising, though the quality of the movie does not change, the perception of its quality varies over time. If the online movie reviews reflect varying perceptions of quality, the parameter estimates for WOM will be spurious. In other words, the WOM may be influenced by other forces, and if both the box office revenues and online ratings are affected by these forces, the parameter estimates will be biased due to the endogeneity problem again. To assess the possible problem of omitted variables, we run model (6) again, but change the WOM valence variable to the averaged amateur rating of reviews posted within one week before weekend *t*. This idea is based on the observation that "cumulative averages" are what the Web site visitors actually see, and the "weekly averages" calculated from a short time span are more representative of the possible changes in the perception of the quality (idiosyncratic changes in WOM). If other forces drive the weekly variations in WOM and online scores, substituting cumulative average scores with weekly average scores should provide more significant parameter estimates for the model. If, however, the robustness check regressions show less significant parameter estimates, they offer evidence that idiosyncratic weekly variations in WOM are not a concern.

Table 7 presents the parameter estimates for this new specification. Again, columns (1), (3), (5), (7) report five-week estimates and the rest report ten-week estimates.

---

[17]For a more elaborated treatment of the issue of seasonality, see Einav (forthcoming).

Table 7. Robustness Check for Idiosyncratic WOM Variation

	Pure Decay		WOM Valence		WOM Variance		Controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Movie quality	-0.1631*	-0.5261***	-0.5529***	-0.7140***	-0.1039	-0.3942***	-0.5251**	-0.5852**
	(0.0947)	(0.0468)	(0.1248)	(0.0802)	(0.2239)	(0.1015)	(0.2354)	(0.1305)
Time decay	-0.0954***	0.0081 **	-0.0895***	0.0092	-0.0980***	0.0037	-0.1327***	-0.0057
	(0.0257)	(0.0073)	(0.0250)	(0.0073)	(0.0253)	(0.0073)	(0.0260)	(0.0091)
WOM valence			0.0411***	0.0204***	0.0193	0.0068	0.0253*	0.0054
(Weekly average)			(0.0098)	(0.0071)	(0.0157)	(0.0074)	(0.0150)	(0.0080)
WOM variance					-0.5363***	-0.4226***	-0.2747*	-0.3536***
					(0.1683)	(0.0838)	(0.1647)	(0.0889)
Competition							-0.0104	-0.0358**
							(0.0236)	(0.0185)
# Theaters							-0.0001***	-0.00005**
							(0.0000)	(0.0000)
Seasonality							0.3094	0.2024
							(0.0537)	(0.0378)
R^2	0.0273	0.0012	0.0563	0.0090	0.0756	0.0326	0.1818	0.0643
F-Test	13.75	1282	14.55	4.76	13.27	11.72	16.33	10.49
Pr>F	0.0002	0.2695	<.0001	.0088	<.0001	<.0001	<.0001	<.0001

Note: Standard errors appear in parentheses.

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

In column (3) and (4), the parameter estimates for WOM valence are significant, but when other variables are added, this is no longer true. It is not surprising that WOM valence in the five-week regressions is generally more significant than in ten-week regressions, because the average score in the first week likely is correlated with the other means of WOM that jumpstart the WOM process. But both the diminishing level of significance and the fading magnitude of the parameter estimates for WOM variance over time suggest that idiosyncratic WOM variation moving along with online ratings is not to be worried.

The second extension involves distinguishing between blockbuster and sleeper movies, which may follow totally different revenue trajectories (and thus have different λ values). Using one model for both types of movies may provide an averaged parameter estimate for the WOM valence.

We estimate models for blockbuster movies and sleeper movies separately. For simplicity, we define a sleeper movie as one whose highest weekly revenue is not obtained in the first weekend[18] . In our sample of 128 movies, 24 are categorized as sleepers. The results of the tests of heteroscedasticity for the blockbuster movies are similar to those for the whole sample, and the standard errors of FGLS estimates are similar to those of OLS. For the sleeper movies, however, some specifications only have fewer than 250 observations and White's test therefore is not reliable for this small sample. MacKinnon and White (1985) raise such concerns about the performance of White's heteroscedasticity consistent covariance matrix (HCCM) in small samples. They refer to the original White's HCCM as HC0, and proposed three more asymptotically equivalent forms of HCCM, HC1 - HC3. With Monte Carlo simulations, Long and Ervin (2000) recommend HC3 for small samples. Therefore, for the sleeper movies, we will use HC3 to derive the robust standard errors[19] .

Table 8. Robustness Check with Blockbusters and Sleepers Separated

	Blockbusters (first 5 weeks)				Sleepers (all periods)			
	Pure Decay	Valence	Variance	Controls	Pure Decay	Valence	Variance	Controls
Movie quality	-0.5041***	-1.0924***	-1.4925***	-1.4381***	0.0245	-0.2755	-0.3086	-0.3332 *
	(0.0628)	(0.1179)	(0.2623)	(0.1816)	(0.0867)	(0.2156)	(0.1434)	(0.1699)
Time decay	-0.0543***	-0.0512***	-0.0493***	0.0233	-0.0333***	-0.0330***	-0.0327***	-0.0299***
	(0.0170)	(0.0163)	(0.0163)	(0.0187)	(0.0075)	(0.0075)	(0.0074)	(0.0064)
WOM valence		0.0631***	0.0918***	0.0534***		0.0491***	0.0511**	0.0435**
		(0.0108)	(0.0199)	(0.0203)		(0.0203)	(0.0214)	(0.0225)
WOM variance			-0.2978*	0.088			0.0381	0.0333
			(0.1745)	(0.1752)			(0.1002)	(0.1327)
Competition				0.0101				-0.0312
				(0.0169)				(0.0319)
# Theaters				0.00015***				0.00002
				(0.0000)				(0.0001)
Seasonality				0.0760**				0.2244***
				(0.0361)				(0.0642)
R^2	0.0256	0.1041	0.1108	0.2374	0.0719	0.0843	0.0846	0.1190
F-Test	10.16	22.36	15.95	17.96	30.85	18.28	12.2	7.88
Pr>F	0.0016	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Note: For blockbusters, standard errors appear in parentheses. For sleepers, the heterogeneity-adjusted HC3 standard errors appear in parentheses.

[18]Note that blockbuster movies are more commonly defined as those with big budgets. They often earn the highest weekly gross during their opening weekends, and their weekly grosses drop sharply over time. Here, we refer to all movies that follow this pattern as blockbusters, so a small production movie that moves quickly out of the cinema might be called a blockbuster.

[19]Briefly, HC0 takes the form $(X'X)^{-1}X'diag[e_i^2]X(X'X)^{-1}$. HC3 is defined as $(X'X)^{-1}X'diag[\ e_i^2/(1-h_{ii})^2\ ]X(X'X)^{-1}$, where $h_{ii}=x_i(X'X)^{-1}x_i'$.

In Table 8, the fixed effect decay parameter for the blockbuster movies is much greater than that for the whole sample; when everything else in the model is controlled for, the average blockbuster movie attracts only 24% of the moviegoers in the previous week. This intrinsic decay parameter is negative for the sleeper movies (except for the pure decay model), but the magnitude is considerably smaller than that of the blockbuster movies. The WOM valence measure remains at the level of 0.04 to 0.05, though it is slightly higher for blockbuster movies in the first five weeks. This result confirms the previous finding, and suggests that a one-point increase in aggregate amateur ratings will bring about 3.9-4.8% more people in the next week. The level of disagreement associated with a movie is not significant for most specifications, which suggests that WOM variance is not an influential variable in the early stage of a movie's life cycle. The seasonality parameter estimate is much greater (and more significant) for sleeper movies, which suggests that in "good" seasons, demand is indeed higher.

That the parameter estimates of WOM valence are almost the same for blockbuster and sleeper movies gives support to our interpretation of this parameter as a measure of the residual impact of WOM valence. To interpret this parameter, both off- and online WOM must be recognized as contributing to the movie-specific decay rates, and there is no way to separate them from the intrinsic quality of a movie. Thus, the parameter estimated is not the impact of online ratings on box office performance but rather the residual impact, that is, the influence not explained by the decay pattern of intrinsic quality. This result likely explains why the parameter estimates are so consistent across blockbuster and sleeper movies, despite their distinctive decay patterns.

Panel Results

The preceding analysis derives an average measure of the rate of decay associated with movies' quality. It also has the advantage of using fewer degrees of freedom. To impute a measure of the intrinsic quality for each movie and to further evaluate the impact of WOM, we reformulate model (6) as a panel model.

The left-hand side can be written as $L_{it} \equiv log\ R_{it}/R_{i,t\text{-}1}$ then we change the time decay parameter η to a fixed effect of time; hence,

$$L_{it} = \lambda_i + \eta_t + \sum_{j=1}^{J} \beta_j x_{jit} + \sum_{k=1}^{K} \gamma_k y_{kit} + \varepsilon_{it} \qquad i = 1, \ldots, N;\ \ t = 1, \ldots, T_i \qquad (7)$$

where movies are indexed with *i*, and weeks after release are indexed with *t*.

Because a movie can be pulled out of the cinema in any week after its release, there are different T_i values for each movie, which makes it an unbalanced panel. From the preceding discussion, it is obvious that λ_i should be treated as a fixed-effects parameter, and its natural interpretation is the quality of the movie *i*. We also expect η_t to be a fixed time effect.

Table 9. Panel Data Results

	Random Two	Fixed One	Fixed Two
Intercept	-0.8883***	-	-
	(0.1336)	-	-
Valence	0.0335***	0.1528**	0.0926***
	(0.0113)	(0.0639)	(0.0312)
Variance	-0.0948	0.0039	-0.0043
	(0.0730)	(0.0839)	(0.0868)
Competition	-0.0273 *	-0.0239	-0.0245
	(0.0149)	(0.0153)	(0.0152)
# Theaters	-0.00005***	-0.00003	-0.00007**
	(0.0000)	(0.0000)	(0.0000)
Seasonality	0.1498***	0.1248***	0.1207***
	(0.0310)	(0.0336)	(0.0342)
R^2	0.0411	0.1882	0.2127

To examine the variance/covariance structure of the data set, we conduct a Hausman specification test (Hausman 1978) to compare the fixed versus random effects models. Under the null hypothesis, individual effects are uncorrelated with other regressors in the model, both OLS and GLS are consistent, but OLS is inefficient. The Hausman test involves comparing the variance covariance matrices of OLS and GLS. If the null hypothesis cannot be rejected, a random effects model is favored. The m-statistic for Hausman test is 10.64 with five degrees of freedom, corresponding to $Pr > m$ = 0.0590. Therefore, the test is inconclusive and indicates a borderline position between the fixed and random effects models.

To confirm the previous findings, we first fit a two-way random effects model using a FGLS procedure, such that the variance components are estimated in the first stage, and then the estimated variance covariance matrix is employed to fit a GLS model. The parameter estimates appear in the first column of Table 9. The result is consistent with that obtained from model (6): The aggregate decay rate is -0.89, which translates to a 58.9% decay across weeks. The parameter estimate for WOM valence is 0.034, with a standard error of 0.0113 (p = .0033). We then generate dummy variables for each movie and run a least square dummy variable regression without intercept. Thus, we can estimate the fixed effects for each movie. The one-way fixed-effects model is highly significant, and the parameter estimate for WOM valence is significant (column 2). Suspecting significant fixed time effects, we also run a two-way fixed-effects model, as reported in the third column. Fixed time effects are significant for the first three weeks. Controlling for these fixed time effects, the parameter estimate for WOM valence is 0.09, which suggests that the WOM residual impact can bring in as much as 9.7% extra revenue. In all three regressions in Table 9, WOM variance is not significant, possibly because with a panel data model formulation, the intercept (in the random effects model) and movie-specific fixed effects (fixed effects models) capture the decay pattern associated with WOM disagreements.

Conclusion

This paper uses weekly movie box office data and Yahoo! Movies review data to estimate the influence of professional and amateur reviews on box office revenues. In the study of the influence of professional reviews, we find that the data set does not support the well-known result of Eliashberg and Shugan (1997). The methodology, adopted from Eliashberg and Shugan, cannot give conclusive results regarding the role of professional reviews any more.

For the amateur reviews, we develop a diffusion model of WOM and exploit the weekly changes in revenue to control for the unobservable intrinsic quality and other time-invariant factors of movies. Using this method, we estimate the residual impact of online amateur reviews on box office revenues. Various robustness tests confirm the soundness of the model.

The results suggest that Yahoo! Movies' online amateur reviews have a positive and statistically significant influence on other people's decision to watch a movie. Specifically, a one-point increase in the grade rating of the aggregate score can induce 4–10% more people to attend the cinemas. The volume and variance of the reviews does not play a significant role in the process, especially when competition, availability of screens, and seasonality are controlled.

In De Vany and Walls (1996)'s model of movie-going, WOM is attributed to be the best explanation for the substantial autocorrelation of growth in demand. However, their model fails to provide an estimate of the magnitude of word of mouth. Our model provides a way to estimate the residual impact of Yahoo!'s online reviews on the box office performance. The diffusion model also allows us to impute a measure of quality for movies. This technique can be used to study the potential impacts of WOM in other online forums.

This paper makes a few simplifying assumptions in an attempt to estimate, with a simplest model, the magnitude of online word of mouth at Yahoo! Movies. One assumption is that, holding the control variables constant, the number of movie-goers in week t is proportional to the number of movie-goers in week $t-1$. This assumption is partially relaxed in model (6) by introducing a time decay variable η. A potential approach we could take is to introduce a movie-specific time-variant parameter to capture the idiosyncratic decay pattern of the movies. That, however, will involve more sophisticated Bayesian methods, and will require a richer data set than what we have now. One can also argue that the movie fixed effect may be changing over time and be different in different genres, and that the time fixed effect may be different across genres, an interesting and promising area of future research would then be examining the potential moderators of the diffusion process.

References

Ainslie, A., Dreze, X., and Zufryden, F.: 2005, Modeling movie lifecycles and market share, *Marketing Science* **24**(3), 508–517.

Akerlof, G.: 1970, The market for lemons: Quality uncertainty and the market mechanism, *Quarterly Journal of Economics* **89**, 488–500.

Basuroy, S., Boatwright, P. H., and Kamakura, W.: 2003, Reviewing the reviewers: The impact of individual film critics on box office performance, Working Paper.

Breusch, T. and Pagan, A.: 1979, A simple test for heteroscedasticity and random coefficient variation, *Econometrica* **47**, 1287–1294.

Chen, Y. and Xie, J.: 2005, Third-party product review and firm marketing strategy, *Marketing Science* 24, 218.240.

Chevalier, J. A. and Mayzlin, D.: 2005, The effect of word of mouth on sales: Online book reviews, Working Paper.

Dellarocas, C., Awad, N., and Zhang, X. M.: 2005, Using online ratings as a proxy of word-of-mouth in motion picture revenue forecasting, Working Paper.

De Vany, A., and Walls, W.D., 1996. Bose-Einstein dynamics and adaptive contracting in the motion picture industry. *Economic Journal*, **106**, 1493-1514.

Dowling, G. R. and Staelin, R.: 1994, A model of perceived risk and intended risk-handling activity, *Journal of Consumer Research* **21**, 119–134.

Dubois, P. and Nauges, C.: 2006, Identifying the effect of unobserved quality and experts' reviews in the pricing of experience goods: Empirical application on Bordeaux wine, Working Paper.

Einav, L.: forthcoming, Seasonality in the U.S. motion picture industry, *RAND Journal of Economics*.

Elberse, A. and Anand, B.: 2005, Advertising and expectations: The effectiveness of pre-release advertising for motion pictures, Working Paper.

Elberse, A. and Eliashberg, J.: 2003, Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures, *Marketing Science* **22**(3), 329–354.

Eliashberg, J., Jonker, J.-J., Sawhney, M. S. and Wierenga, B.: 2000, Moviemod: An implementable decision-support system for prerelease market evaluation of motion pictures, *Marketing Science* **19**, 226–243.

Eliashberg, J. and Shugan, S. M.: 1997, Film critics: Influencers or predictors? *Journal of Marketing* **61**, 68–78.

Godes, D. and Mayzlin, D.: 2004, Using online conversations to study word-of-mouth communication, *Marketing Science* **23**(4), 545–560.

Hausman, J. A.: 1978, Specification tests in econometrics, *Econometrica* **46**, 1251–1271.

Holbrook, M. B.: 1999, Popular appeal versus expert judgments of motion pictures, *Journal of Consumer Research* **26**(2), 144–155.

Jedidi, K., Krider, R. E., and Weinberg, C. B.: 1998, Clustering at the movies, *Marketing Letters* **9**(4), 393–405.

Levinsohn, J. and Petrin, A.: 2003, Estimating production functions using inputs to control for unobservables, *Review of Economic Studies* **70**(243), 317–341.

Litman, B. R.: 1983, Predicting the success of theatrical movies: An empirical study, *Journal of Popular Culture* **17**, 159–175.

Litman, B. and Kohl, L.: 1989, Predicting financial success of motion pictures: The 80's experience, *Journal of Media Economics* **2**, 35–50.

Long, J. and Ervin, L.: 2000, Using heteroscedasticity-consistent standard errors in the linear regression model, *The American Statistician* **54**, 217–224.

Lopate, P. (ed.): 2006, American Movie Critics: An anthology from the silents until now, Library of America.

MacKinnon, J. and White, H.: 1985, Some heteroscedasticity-consistent covariance matrix estimators with improved finite sample properties, *Journal of Econometrics* **29**, 305–325.

Reinstein, D. A. and Snyder, C. M.: 2005, The influence of expert reviews on consumer demand for experience goods: A case study of movie critics, *Journal of Industrial Economics* **LIII**(1), 27–51.

Sawhney, M. and Eliashberg, J.: 1996, A parsimonious model for forecasting gross box office revenues of motion pictures, *Marketing Science* **15**(2), 321–340.

Sochay, S.: 1994, Predicting the performance of motion pictures, *Journal of Media Economics* **7**, 1–20.

West, P. M. and Broniarczyk, S. M.: 1998, Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus, *Journal of Consumer Research* **25**(1), 38–51.

White, H.: 1980, A heteroskecastic-consistent covariance matrix estimator and a direct test of heteroskedasticity, *Econometrica* **48**, 817–838.

Zufryden, F.: 1996, Linking advertising to box office performance of new film releases: A marketing planning model, *Journal of Advertising Research*, 29–41.

Zufryden, F.: 2000, New film Web site promotion and box-office performance, *Journal of Advertising Research*, 55–64.

Appendix: Description of Major Empirical Variables

Movie Quality: λ – A derived measure from Model (6).

Time Decay: η – Estimates the speed of decreasing viewers over time.

WOM Valence: an average score calculated from the reviewers' ratings

WOM Variance: The inverse of the coefficient of variation

Competition: Number of new movies, number of movies in the same genre

Theaters: Number of theaters showing the movie

Seasonality: An index calculated from 3 years of weekly box office data, it measures the seasonal phenomenon that some weeks (July the 4th, Labor Day weekend, etc.) tend to be more crowded than others.